

# THE CAMBRIDGE STATISTICS DISCUSSION GROUP

**Wednesday 14th March 2007 7:15 for 7:45**

Small (Tom Ap Rees) Lecture Theatre, Department of Plant Sciences,  
Downing Site, Cambridge

## **Retrieving Information using a Bayesian Model of Generalization**

**Zoubin Ghahramani**

Dept of Engineering  
Cambridge

**Abstract:** Humans readily learn new concepts after observing a few examples and show extremely good generalization to new instances. In contrast, search tools on the internet exhibit little or no learning and generalization. Here we present a new framework for retrieving information based on principles governing how humans learn new concepts and generalize. Given a query consisting of a set of items representing some concept, our method automatically infers which other items are relevant to that concept and retrieves them. Unlike previous such tools, which are based on simple text queries, our method leverages the rich and subtle information provided by a query consisting of a set of items. Moreover our work shows that retrieval can be firmly grounded in a Bayesian statistical model of human learning and generalization. Finally, the underlying computations reduce to an extremely efficient sparse linear equation, making it practical for large scale retrieval problems. We show five example applications including searches for scientific articles, images, proteins, and movies. Joint work with Katherine A Heller, University College London

**Speaker:** Zoubin Ghahramani is Professor of Information Engineering at the University of Cambridge, UK. He obtained BA and BSE degrees from the University of Pennsylvania, and a PhD in 1995 from MIT working with Prof Mike Jordan. He was a Postdoctoral Fellow in Computer Science at the University of Toronto working with Prof Geoff Hinton.

His work has included research on human sensorimotor control, cognitive science, statistics, and machine learning. His current focus is on Bayesian approaches to statistical machine learning, with applications to bioinformatics, information retrieval, and other areas. He has published over 100 peer reviewed papers, and serves on the editorial boards of several leading journals in the field, including JMLR, JAIR, IEEE PAMI, Machine Learning, and Bayesian Analysis. He also serves on the Board of the International Machine Learning Society. Zoubin Ghahramani is also Associate Research Professor of Machine Learning at Carnegie Mellon University.

**Directions:** The Department of Plant Sciences is located on the Downing site situated between Downing Street and Tennis Court Road (see map below). Go through the second door you come to in the building immediately on your left as you enter from Tennis Court Road. Once inside the building follow the arrows. Refreshments will be served in the first floor kitchen. It is not possible to park on the Downing site in the evening. The nearest car park is at Lion Yard. Arrivals after 7:45pm can gain admittance by contacting the secretary on 07761769436.

### **Provisional Next Meetings:**

4th April – Sue Welham (Rothamsted Research).

3rd May – Brian Tom (Biostatistics).

9th October – Matt Whiley (Amgen).

**Supper:** Some members eat regularly in the University Centre before each meeting at **5-45pm**. Feel free to join them.

**Subscriptions:** of 1 pound are now due for attending the 2006-2007 session.

**Secretary:** Peter Watson, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF;

telephone 01223 355294 Extension 801; E-mail [peter.watson@mrc-cbu.cam.ac.uk](mailto:peter.watson@mrc-cbu.cam.ac.uk)

PLANT SCIENCES

